PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68, G01N 33/53

A1

(11) International Publication Number: WO 96/23903

(43) International Publication Date: 8 August 1996 (08.08.96)

(21) International Application Number: PCT/SE96/00095

(22) International Filing Date: 30 January 1996 (30.01.96)

(30) Priority Data:
9500341-4 30 January 1995 (30.01.95) SE

(71)(72) Applicant and Inventor: LANDEGREN, Ulf [SE/SE]; Eksoppsvägen 16, S-756 46 Uppsala (SE).

(72) Inventor; and
(75) Inventor/Applicant (*for US only*): LAGERQVIST, Arild [SE/SE]; Apelbergsgatan 54, S-111 37 Stockholm (SE).

(74) Agents: WIDÉN, Björn et al.; Pharmacia AB, Patent Dept., S-751 82 Uppsala (SE).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD FOR DETECTING NUCLEIC ACID SEQUENCE VARIATIONS

(57) Abstract

A method of analyzing nucleic acid-containing samples for sequence variations relative to standard nucleic acid sequences, which method comprises providing a single-stranded standard nucleic acid sequence immobilized on a solid support, hybridizing a nucleic acid strand derived from a sample to the immobilized strand, subjecting the nucleic acid complex formed to (i) mismatch-induced cleavage or (ii) mismatch-terminated extension reactions, and detecting possible cleavage or extension-termination by optical measurement on the solid support.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia
AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

METHOD FOR DETECTING NUCLEIC ACID SEQUENCE VARIATIONS

The present invention relates to a method for detecting variations in the sequences of nucleic acid fragments, particularly in the DNA sequences of genes or gene fragments in patient samples in relation to the wild type genes.

Clinical analyses of DNA sequences are typically directed to determining how a gene in a patient sample differs from a prototypical normal sequence. DNA sequencing through the chain termination method developed by Sanger and Coulson (Sanger et al., Proc. Natl. Acad. Sci. USA 1977; 74: 5463-5467), and the chemical degradation method developed by Maxam and Gilbert (Maxam and Gilbert, Proc. Natl. Acad. Sci. USA 1977; 74: 560-564), or using techniques such as Sequencing By Hybridization (SBH) or Sequencing By Synthesis (see e.g. WO 93/21340) all have the potential to identify mutations and in the same process also reveal the consequence of the mutation at the level of protein coding etc.

For screening purposes, however, it is often sufficient, at least initially, to identify deviations from the normal sequence but without directly revealing how a sequence differs from the normal one or only roughly locating the mutation. There are a number of such techniques which speed up analysis as compared to those that involve DNA sequence determination.

Methods to scan or screen for mutations may be divided into two groups, i.e. those that identify mutations trough altered properties of heteroduplexes, i.e. base-paired molecules composed of one strand from the normal sequence and a complementary strand derived from the patient sample, and those that observe properties of single stranded molecules or of homoduplexes. Examples from the first category of methods are RNAse cleavage of mismatched positions in hybrids between an RNA strand and a complementary DNA strand (Myers R. M. et al., Science 1985; 230:1242-1246). It is also possible to detect mismatches in

heteroduplexes via their effect upon the melting behaviour of the molecules as they migrate in a gel under increasingly denaturing conditions (Myers R. M. et al., Nature 1985; 313: 495-498). One of the more popular methods uses base-modifying chemistry to selectively sensitize mismatched bases for subsequent cleavage (Cotton R. G. H. et al., Proc. Natl. Acad. Sci. USA 1988; 85: 4397-4401; and Montadon A. J. et al., Nucl. Acids. Res. 1989; 1 (9): 3347-3358). There is also a method where mismatched bases are modified so that the modified positions in a replication template will terminate a subsequent primer-extension reaction (Ganguly A., Prockop D. J., Nucl. Acids Res. 1990; 18 (13): 3933-3939). Recently, enzyme systems serving the purpose to detect mismatched bases in DNA duplexes have been applied for this purpose (Lu A-L, Hsu I-C, Genomics 1992; 14: 249-255; and Yeh Y-C et al., J. Biol. Chem. 1991; 266: 6480-6484). Such enzyme systems may cleave most or all mismatched positions in DNA strands with a length of at least several hundred bases. An exemplary such enzyme system is T4 endonuclease VII (Youil R. et al, Proc. Natl. Acad. Sci. USA 1995; 92: 87-91).

Typically, in the methods where sequence differences are demonstrated through the cleavage or modification of mismatched positions in heteroduplexes, the results are evaluated by gel electrophoretic separation of the strands, providing an estimate of the position of the mismatch. Recently, automated sequencers equipped for fluorescent detection of the molecules have been used for this purpose (Verpy E. et al., Proc. Natl. Acad. Sci. USA 1994; 91: 1873-1877). The necessary electrophoretic separation is, however, laborious and time-consuming.

WO 93/20233 discloses a method for identifying a base pair mismatch at a site in a nucleic acid by labelling a single stranded target nucleic acid sequence at two sites on either side of the target site, fixing the doubly labelled nucleic acid to a solid support at one end, hybridizing a corresponding wild type nucleic acid fragment to the target sequence, exposing the nucleic acid hybrid to

a mismatch-cleaving enzyme, and after a wash detecting the presence of both labels or one label, that on the end fixed to the solid support or that washed away. This method avoids electrophoretic separation, but is disadvantagous in that it inter alia does not permit use for the simultaneous screening of a plurality of different target sequences on a single solid support.

The object of the present invention is to provide an improved method for identifying sequence discrepancies between nucleic acid sequences, such as normal and patient sample genes or gene fragments, by mismatch-techniques, like those listed above, but which does not use any gel electrophoretic separation, and which may readily be adapted to array formats for multiple screening purposes.

According to the invention, this object is achieved by providing a single-stranded prototypical normal nucleic acid sequence, or standard (wild type) sequence, immobilized on a solid support, hybridizing a nucleic acid strand derived from a patient sample to the immobilized strand, subjecting the nucleic acid hybridization complex formed to (i) mismatch-induced cleavage or (ii) mismatch-restricted extension reactions, and detecting possible cleavage or extension-termination by optical measurement on the solid support.

The nucleic acid sequences are preferably DNA sequences, such as genomic DNA sequences.

In a preferred embodiment, sets of the same or different normal nucleic acid strands, especially DNA, are immobilized in a linear or, preferably, 2-dimensional planar array to permit either several patient derived samples to be tested in parallel, or, more preferably, several nucleic acid sequences, especially gene fragments, derived from one individual to be tested at the same time.

When carrying out such an embodiment of the invention by mismatch-induced cleavage of DNA hybrids, either the free ends of the patient DNA strands and/or the normal strands may be labelled, e.g. with a dye, such as with a fluorophore or a chromophore. After washes, the number of

DNA strands bound to their respective positions may be estimated by measuring the local label signal, e.g. fluorescence. Then, cleavage of DNA heteroduplexes at mutation positions is performed using any of the above-mentioned techniques, followed by washes, to remove single strands or base-paired DNA segments, depending on whether the molecules have been cleaved in one or both strands. In general, it is thus desirable to attach the molecules to the support so that the strands that bear detectable functions are not removed under denaturing conditions, unless they have been cleaved. Examples of linkages of the DNA to the support that are suitable include binding by a biotin-avidin/streptavidin interaction or covalent bonds, e.g. formed by chemically coupling DNA to the support or by ligating the DNA strand to an oligonucleotide, stably bound to the support. Techniques suitable for such attachment are known to those skilled in the art. After cleavage and denaturing washes, another measurement of the local fluorescence in each position is performed, and the ratio of fluorescence after versus before cleavage is estimated. Any significant reduction of the fluorescence as compared to that before cleavage indicates mismatches in at least some of the strands from the patient samples.

An analogous procedure may be used in the case of mismatch-restricted extension, where extension products may be labelled by incorporating detectable functions as modified nucleotides during the extension reaction.

Measurement of cleavage or extension termination may also be performed by optical "label-free" techniques, such as, for example, mass or refractive index sensing techniques based on evanescent wave sensing, such as surface plasmon resonance (SPR) based methods.

Differences in the tendency to non-specific cleavage as a function of factors such as hybrid length, base composition or curvature may be weighted in as a background against which to compare the results of the analysis. The above method permits a very large number of templates to be simultaneously analyzed in the described manner, and also

relatively small contributions of mutant sequences may give rise to a detectably different signal.

Complications of the analysis due to the presence of polymorphisms, i.e. normal variations in a sequence under study, may be overcome by investigating several independent sequence variants.

There are a number of techniques known in the art for immobilizing the desired templates to the solid support. For example, short oligonucleotides arrayed on a two-dimensional support may be used to ligate one strand from a specific PCR product by using primers that provide a 5' single stranded extension, as described by Newton C. R., Nucl. Acids Res. 1993; 21: 1155-1162. This technique would permit the assortment of large sets of different PCR products to the appropriate positions in the array by hybridization to the corresponding immobilized oligonucleotides via a splint.

The arrays themselves supporting oligonucleotides or longer nucleic acid strands, such as DNA strands, in defined positions and to which the templates may be immobilized may be prepared by any of a number of techniques known in the art. One such technique, which is described in the International patent application No. PCT/SE95/01420, involves the bundling of structures containing DNA strands, followed by sectioning and deposition on a planar surface. Alternatively, the oligonucleotides may be bound to the solid support via a specific binding pair, such as biotin and avidin or streptavidin. For example, the primers can be provided with biotin handles in connection with their preparation, and then the biotin-labelled oligonucleotides can be attached to a streptavidin-coated support. The oligonucleotides can also be bound by a linker arm, such as a covalently bonded hydrocarbon chain, e.g. a C_{10-20} chain. As another alternative, the oligonucleotides can be bound directly to the solid support, such as by epoxide/amine coupling chemistry.

The solid support can be a plate or chip of glass, silicon or other material. The solid support can also be coated, such as with gold or silver. Coating may facilitate attachment of the oligonucleotides to the solid support.

Measurement of the fluorescence from fluorophore-labelled duplexes may be performed by methods known in the art, such as by a two-dimensional diode array or a CCD (charge-coupled device) camera, for example.

The patient samples may be simply labelled by selecting a 5' fluorophore-labelled primer in the amplification reaction used for amplifying the DNA strand or strands of interest. It is also possible to label the standard sequence on the solid support in the 3' position by tailing with terminating fluorophore-labelled nucleotides (Prober J. M. et al., Science 1987; 238: 336-341), and using the enzyme terminal deoxynucleotidyl transferase (Maniatis T. et al., Molecular cloning: A laboratory manual. New York: Cold Spring Harbor Press, 1982).

The hybridization of patient strands to the arrayed templates is simplified if these are first rendered single-stranded, e.g. by digestion with lambda exonuclease (Higuchi R. G. et al., Nucl. Acids Res. 1989; 17(14): 1989, and Nikiforov T. T. et al., PCR Meth. Applic. 1994; 3: 285-291). In this regard the fluorophore label, which is present at the 5' end of the amplified strands, conveniently protects this strand against degradation.

Among advantages of the method of the invention may be mentioned the circumstance that a normal sequence may be selected as a standard against which to compare the patient sample, which means that both homozygous and heterozygous mutations may be monitored. The technique may have obvious advantages for scanning for clinically important mutations, ultimately in all estimated human 65,000 genes. However, the technique could also be useful as a forensic tool, rapidly identifying differences between DNA samples or for the typing of genes such as transplantation genes. In genetic linkage analysis, the method of the invention would

provide access to a much larger set of genetic markers than RPLFs or microsatellites, since any point mutation occurring in segments of DNA could be scored. In particular, the method of the invention would be highly useful to identify the location of homozygous genomic regions in individuals affected by recessive disorders by being homozygous by decent, i.e. having inherited the same mutated gene through parental lineages.

The invention will be illustrated further by the following non-limiting example.

EXAMPLE

A PCR product from an individual having the normal form of the amplified globin gene is generated in a reaction where one of the primers has a 5' extension-sequence, interrupted by non-nucleotide residues, such as for instance hexaethylene glycol residues (HEG) (other oligonucleotide sequence modifications serving a similar purpose have been described by Newton et al., Nucl. Acids Res. 1993; 21:1155-1162). After PCR, the 5' extension will remain single stranded and is used to hybridize and ligate the PCR product to a solid support which has previously been modified by the coupling of a suitable oligonucleotide with a free 3′end, and to which a complementary oligonucleotide has been hybridized such that this hybridized oligonucleotide can also hybridize to the 5' extension of the PCR product, and permit the PCR product to be ligated to the oligonucleotide on the support. After denaturing washes, the free 3' end of the amplified strand remaining on the support is modified with a dideoxynucleotide with an added fluorophore, using the enzyme terminal deoxynucleotide transferase. A corresponding PCR product is derived from a patient in order to investigate if the globin gene in this patient differs from the normal sequence. In this amplification reaction, a 5' phosphorylated primer lacking any non-nucleotidic sequences is used instead of the HEG-modified primer used for the amplification of the normal gene. Instead the opposite primer is modified through the

addition of a biotin 5' reside which will protect this strand from digestion by the 5' exonuclease λ-exonuclease (another 5' modification, serving a similar purpose has been described by Nikiforov et al., PCR Meth. Appl. 1994; 3:285-291). An excess of the single strands from the amplified patient sample is then hybridized to the single strands bound to the support. After the fluorescence from the immobilized molecules has been recorded, the duplex molecules are exposed to reagents that cleave mismatched positions in duplex DNA, such as T4 endonuclease VII, as taught by Youil et al., Proc. Natl. Acad. Sci. USA 1995; 92: 87-91. After denaturing washes, another fluorescence reading is then taken to determine if the support-bound strands have undergone cleavage, indicative of a mismatch in the hybrid with the strand derived from the patient. Any such reduction then indicates a sequence variation in the globin gene sequence of the patient and may prompt further analysis of the gene in this patient, e.g. by DNA sequence analysis.

The above described example of carrying out the invention may be modified in two ways to increase the probability and speed of detecting any mutations. The supports may be designed so that not only the 5' end of the standard sequence but also the 3' end of the patient sample are stably attached to the support, e.g. through ligation, and, along with the 3' end of the standard sequence, also the 5' end of the patient sample may be modified with a fluorophore, for instance by attaching a fluorophore instead of the biotin group in the example above. In this manner, cleavage of either or both strands of the immobilized heteroduplex may be detected, increasing the probability of detecting mutations. The other suggested modification of the above protocol is by performing the analysis for a large number of samples in parallel. Thus, at defined locations on a 2-dimensional array the corresponding genes or gene fragments from many patients may be hybridized. Alternatively, and more importantly, many different gene sequences in one individual may be

compared to standard variants of the corresponding sequences, immobilized in discrete locations.

The invention is, of course, not restricted to the embodiments specifically described above, but many changes and modifications may be made within the scope of the general inventive concept as defined in the following claims.

CLAIMS

1. A method of analyzing nucleic acid-containing samples for sequence variations relative to standard nucleic acid sequences, which method comprises providing a single-stranded standard nucleic acid sequence immobilized on a solid support, hybridizing a nucleic acid strand derived from a sample to the immobilized strand, subjecting the nucleic acid complex formed to (i) mismatch-induced cleavage or (ii) mismatch-terminated extension reactions, and detecting possible cleavage or extension-termination by optical measurement on the solid support.

2. The method according to claim 1, wherein said nucleic acid is DNA.

3. The method according to claim 1 or 2, wherein said standard sequence is a prototypical human gene sequence or a part thereof.

4. The method according to claim 1, 2 or 3, wherein said mismatch-induced cleavage comprises sensitizing mismatched bases by base-modifying chemical agents and subsequently cleaving the nucleic acid complex at the mismatch site by chemical means.

5. The method according to any one of claims 1 to 4, wherein said mismatch-induced cleavage comprises cleaving the nucleic acid complex at the mismatch site by means of an enzyme system, such as, for example, T4 endonuclease VII.

6. The method according to any one of claims 1 to 5, wherein fluorescence from a fluorophore label or luminescence from a luminescer label is measured.

7. The method according to any one of claims 4 to 6, wherein the free ends of nucleic acid strands derived from the sample are labelled.

8. The method according to any one of claims 4 to 7, wherein the free ends of the immobilized standard nucleic acid strands are labelled.

9. The method according to any one of claims 1 to 8, wherein said optical measurements are performed prior to and after said cleavage or chain extension reactions and the measurement results are compared with each other.

10. The method according to any one of claims 1 to 9, wherein sets of different standard nucleic acid sequences are immobilized at defined positions on the solid support and that a number of different nucleic acid fragments thereof derived from one individual are analyzed simultaneously.

11. The method according to claim 10, wherein said sets of standard nucleic acid sequences are provided in a two-dimensional array.

12. The method according to any one of claims 1 to 11 for screening of gene fragments for DNA sequence variations with respect to the corresponding wild type DNA sequences, which method comprises immobilizing a set of wild type DNA strands in an array on a solid support, hybridizing different gene fragments derived from one individual to said immobilized wild type strands, subjecting the DNA hybrids formed to (i) mismatch-induced cleavage or (ii) mismatch-terminated extension reactions, and detecting by optical measurement on the solid support cleavage or extension-termination of one or more of the DNA hybrids as indicative of mismatch between the gene fragment and the corresponding wild type sequence.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 96/00095

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12Q 1/68, G01N 33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12Q, G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, PCI, MEDLINE, BIOSIS, DERWENT BIOTECH, ABS. EMBASE, CLAIMS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
D,X	WO 9320233 A1 (UNIVERSITY OF MARYLAND AT BALTIMORE), 14 October 1993 (14.10.93), page 11, line 15 - page 12, line 15; page 13, line 16 - line 27; page 24, line 5 - line 28 --	1-12
A	Nature, Volume 353, October 1991, R.G.H. Cotton et al, "Mutation detection" page 582 --	1-12
A	WO 9322457 A1 (MASSACHUSETTS INSTITUTE OF TECHNOLOGY), 11 November 1993 (11.11.93), page 10, line 9 - line 24 -- --------	1-12

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" ertier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed
- "T" later document published after the international filing date or pri date and not in conflict with the application but cited to understa the principle or theory underlying the invention
- "X" document of particular relevance: the claimed invention cannot b considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance: the claimed invention cannot t considered to involve an inventive step when the document is combined with one or more other such documents, such combin being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 April 1996	02 -05- 1996
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Patrick Andersson Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

Information on patent family members

01/04/

International application No.
PCT/SE 96/00095

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A1- 9320233	14/10/93	NONE	
WO-A1- 9322457	11/11/93	NONE	